



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **06348811 A**(43) Date of publication of application: **22.12.94**

(51) Int. Cl. **G06F 15/62**
G09G 5/00
G10L 3/00
G10L 3/00

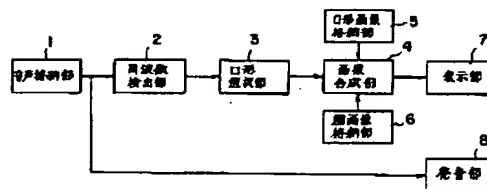
(21) Application number: **05135755**(71) Applicant: **SHARP CORP**(22) Date of filing: **07.06.93**(72) Inventor: **SAKAMOTO KENJI****(54) MOVING IMAGE DISPLAY DEVICE****(57) Abstract:**

PURPOSE: To accurately identify an input voice in a short time and to obtain a face image provided with accurate mouth shape synchronizing with the input voice by deciding the mouth shape of a vocalized voice based on a sampled frequency component, and obtaining a synthesized face image by synthesizing a mouth shape image with the mouth area of a human face image.

CONSTITUTION: A frequency detecting part 2 extracts frequencies of high-pass and low-pass areas in an audio signal inputted from a voice storage part 1. A mouth shape selection part 3 decides the mouth shape to be displayed based on the detection result of the frequency component of the frequency detecting part 2. An image synthesizing part 4 reads out image data in accordance with the mouth shape selected at the mouth shape selection part 3 from a mouth shape image storage part 5, and simultaneously, reads out the image data representing the human face image from a face image storage part 6, and forms the synthesized face image by fitting the mouth shape image in the mouth area of the face image. Furthermore, a display part 7 outputs the synthesized image formed at the image synthesizing part

4 by the control of a synchronous control part synchronizing with voice output from a vocalizing part 8.

COPYRIGHT: (C)1994,JPO



(51) Int.Cl. ⁵	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 15/62	3 4 0 D	8125-5L		
G 0 9 G 5/00	A	8121-5G		
G 1 0 L 3/00	S	8946-5H		
	5 5 1 G	9379-5H		

審査請求 未請求 請求項の数 2 O L (全 4 頁)

(21) 出願番号 特願平5-135755

(22) 出願日 平成5年(1993)6月7日

(71) 出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 坂本 憲治

大阪府大阪市阿倍野区長池町22番22号 シ

ャープ株式会社内

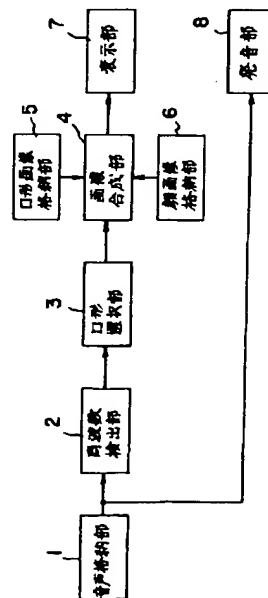
(74) 代理人 弁理士 藤本 博光

(54) 【発明の名称】 動画像表示装置

(57) 【要約】

【目的】 人間の表情を人工的に合成して表示装置の画面上に表示する動画像表示装置に関し、入力音声を経時間で正確に識別して入力音声と同期した口形画像を表示することを目的とする。

【構成】 発声する音声の低域および高域の周波数成分を抽出する周波数検出部と、周波数検出部で抽出した周波数成分に基づいて発声する音声の口形を決定する口形選択部と、この決定した口形に応じた口形画像を人間の顔画像の口領域に合成して合成顔画像を得る画像合成部と、音声の発声に同期して合成顔画像を表示する表示部とから構成する。



【特許請求の範囲】

【請求項1】 発声する音声の低域および高域の周波数成分を抽出する周波数検出部と、
前記周波数検出部で抽出した周波数成分に基づき前記発声する音声の口形を決定する口形選択部と、
前記決定した口形に応じた口形画像を人間の顔画像の口領域に合成して合成顔画像を得る画像合成部と、
前記音声の発声に同期して前記合成顔画像を表示する表示部と、を備えることを特徴とする動画像表示装置。

【請求項2】 前記周波数検出部は、前記発声する音声の低域を検出する第1の帯域フィルタと前記発声する音声の高域を検出する第2の帯域フィルタとを備え、前記口形選択部は前記第1および第2の帯域フィルタの両出力値に基づいて前記発声する音声の口形を決定することを特徴とする請求項1記載の動画像表示装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、人間の表情を人工的に合成して表示装置の画面上に表示することにより、オペレータである人間と機械とのコミュニケーションを円滑に行えるようにした動画像表示装置に関する。

【0002】

【従来の技術】従来、この種の表示装置としては、発声する内容をテキストにし、このテキストに応じて発声する音声の口形と発声時間とを決定し、この決定した口形の画像を人間の顔画像の口領域に合成し、この合成顔画像をテキストから合成した音声と同期して表示する装置がある。

【0003】また、音声認識技術を用いて入力音声の母音を抽出し、この抽出した母音に応じた口形を決定し、この決定した口形の画像を人間の顔画像の口領域に合成し、この合成顔画像を入力音声の発声速度に同期して表示する装置がある（原島ら；「表情アニメーション作成のためのシナリオ記述ツールとリアルタイム動画像表示」；信学技法HC91-57, P23~30参照）。

【0004】

【発明が解決しようとする課題】ところが、前述した2種類の装置のうち、第1の装置は、テキストに応じて口形を決定するため口形は正確に表現できるが、音声はテキストから合成するため自然性に欠けるといった不都合がある。

【0005】また、第2の装置は、音声認識技術によって入力音声をテキスト化してから口形を決定するため、音声は正確に表現できるが、口形は音声の認識を誤ると不自然なものが選択され、正確な合成顔画像が得られないといった不都合がある。また、この装置は処理に時間がかかり過ぎるといった不都合もある。

【0006】そこで、本発明は、入力音声を短時間で正確に識別し、入力音声と同期して正確な口形を有する顔画像を表示することを目的とする。

【0007】

【課題を解決するための手段】本発明による動画像表示装置は、発声する音声の低域および高域の周波数成分を抽出する周波数検出部と、周波数検出部で抽出した周波数成分に基づき発声する音声の口形を決定する口形選択部と、この決定した口形に応じた口形画像を人間の顔画像の口領域に合成して合成顔画像を得る画像合成部と、音声の発声に同期して合成顔画像を表示する表示部とから構成する。

【0008】この場合、周波数検出部は、発声する音声の低域を検出する第1の帯域フィルタと高域を検出する第2の帯域フィルタとを備え、口形選択部はこの第1および第2の帯域フィルタの両出力値に基づいて発声する音声の口形を決定する。

【0009】

【作用】本発明の構成において、周波数検出部は発声する音声の低域および高域の周波数成分を抽出し、口形選択部はこの抽出した周波数成分に基づいて口形を決定する。口形を決定するうえで重要なのは調音位置であるが、母音の調音位置は主に第1フォルマントおよび第2フォルマントの位置によって決まる。

【0010】このため周波数検出部では低域側の第1の帯域フィルタによって第1フォルマントを抽出し、高域側の第2の帯域フィルタによって第2フォルマントを抽出する。口形選択部はこの2種類の帯域フィルタで抽出した周波数成分の割合から入力音声を摩擦子音、パスバ、5種類の母音および無音に大別し、それぞれに応じた口形を決定する。合成画像部はこうして決定した口形に対する口形画像を顔画像の口領域に合成して合成顔画像を得、表示部はこの合成画像を音声の発声に同期して表示する。

【0011】

【実施例】図1は、本発明による動画像表示装置の一実施例を示すブロック図である。本実施例において、音声格納部1は自然音声を格納するためのもので、予め収集した音声であっても、リアルタイムで入力した音声であってもよい。

【0012】周波数検出部2は音声格納部1から入力された音声信号の低域および高域の周波数成分を抽出するためのもので、低域側の帯域フィルタBPF1と高域側の帯域フィルタBPF2とによって構成されている。この2つの帯域フィルタBPF1およびBPF2は、図2に示すように、周波数1200Hzを境界とする周波数特性を有しており、この境界周波数（1200Hz）は母音の第1ホルマントと第2ホルマントとを区別しやすくするために設定した値である。

【0013】口形選択部3は周波数検出部2における周波数成分の検出結果に基づいて、表示すべき口形を決定するためのもので、前述した2つの帯域フィルタBPF1およびBPF2の出力値から口形を決定する。図3

は、口形選択部3における口形決定のための特性図で、横軸に帯域フィルタBPF1の出力値、縦軸に帯域フィルタBPF2の出力値を示し、この2つの出力値の相関から口形を決定する。

【0014】図中、fr, rel, a, i, u, e, o, mmは、それぞれ口形を表す記号で、frは/s/（サ行音）のような摩擦性子音を発声するときの口形、relは発声していないときの口形、aは母音/a/を発声するときの口形、iは母音/i/を発声するときの口形、uは母音/u/を発声するときの口形、oは母音/o/を発声するときの口形、mmは/ma/を発声するときのような唇を閉じたときの口形である。

【0015】従って、例えば帯域フィルタBPF1の出力レベルがa1以下で、かつ帯域フィルタBPF2の出力レベルがb1以上であれば、口形選択部3は記号frの口形を選択する。

【0016】画像合成部4は口形選択部3で選択した口形に対応する画像データを口形画像格納部5から読み出し、同時に顔画像格納部6から人間の顔画像を表示する画像データを読み出して、図4に示すように、顔画像Fの口領域Mに、口形画像を嵌め込んで合成顔画像を形成する。口形画像格納部5には、図5に示すように、前述した8種類の口形fr, rel, a, i, u, e, o, mmに対応する口形の画像データが格納されている。

【0017】表示部7は画像合成部4で形成した合成顔画像を、同期制御部（不図示）からの制御によって、発音部8から発声する音声出力と同期して表示するもので

ある。この場合、顔画像に嵌め込む口形画像の切り換えを、口形が変化したときのみ実行するようにすれば、処理量を大幅に削減することができる。

【0018】

【発明の効果】本発明によれば、簡易な手段で正確かつ迅速に入力音声の口形を決定することができ、その結果、音声と同期した違和感のない口形を有する合成顔画像を表示することができるので、オペレータと機械との間でのヒューマン・インターフェースが可能となる。

【図面の簡単な説明】

【図1】本発明による動画像表示装置の一実施例を示すブロック図である。

【図2】図1に示す周波数検出部の周波数特性を示す図である。

【図3】図1に示す口形選択部の特性図である。

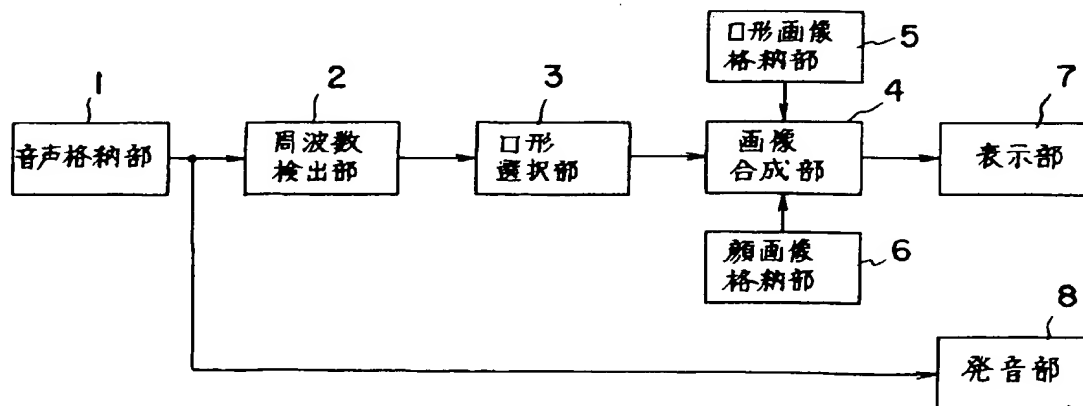
【図4】顔画像と口領域とを表す図である。

【図5】口形画像の一例を示す図である。

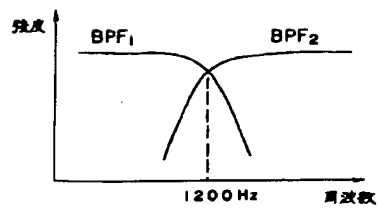
【符号の説明】

- 1 音声格納部
- 2 周波数検出部
- 3 口形選択部
- 4 画像合成部
- 5 口形画像格納部
- 6 顔画像格納部
- 7 表示部
- 8 発音部

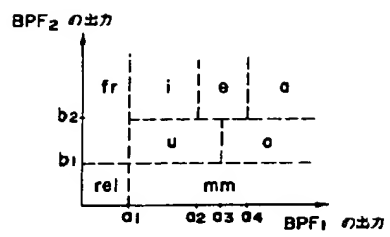
【図1】



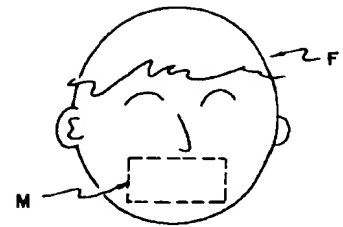
【図2】



【図3】



【図4】



【図5】

